DOCUMENT RESUME

ED 409 870 IR 018 459

AUTHOR Sheehan, Joseph; Tessmer, Martin
TITLE A Construct Validation of the Mental Models Learning Outcome Using Exploratory Factor Analysis.
PUB DATE 97
NOTE 19p.; In: Proceedings of Selected Research and Development Presentations at the 1997 National Convention of the Association for Educational Communications and Technology (19th, Albuquerque, NM, February 14-18, 1997); see IR 018 421.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Construct Validity; *Factor Analysis; Factor Structure; Higher Education; *Knowledge Representation; Learning Strategies; *Outcomes of Education; Problem Solving; Spreadsheets; Testing; Troubleshooting
IDENTIFIERS *Mental Models

ABSTRACT

A mental model is a knowledge structure composed of concepts and the relations between them. Mental models are distinct from declarative and procedural knowledge--they go beyond semantic relationships and skills acquisition and contain varied intellectual skills and knowledge. This study describes an initial investigation into the construct validity of mental models as indication of a distinct learning outcome. The study considered the relationship between mental models, declarative knowledge, concept learning, problem solving, and troubleshooting performance. College junior-level Operations Management students (n=21) were tested on their knowledge and use of spreadsheets through multiple choice and matching quizzes developed using the Pathfinder Associative Network data collection method. Study limitations included: subject mortality; generalizability; missing variables; insufficient number of subjects; invalidated instruments; and convergent validity. Appendices include: definitions of structured knowledge, construct validity, and factor analysis terms; sample questions from the declarative knowledge quiz; concept matching quiz; and SPSS factor analysis results. (Contains 55 references.) (Author/SWC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 409 870

A Construct Validation of the Mental Models Learning Outcome Using Exploratory Factor Analysis

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Joseph Sheehan
Martin Tessmer
University of South Alabama

Abstract

This study was designed as an initial investigation into the construct validity of mental models. The content area selected was the use of spreadsheets. The intent was to analyze the correlation between verbal information and concept learning with Pathfinder measures of mental models acquisition. Although exploratory procedures were used for the factor analysis, the authors entered the analyses with research-based hypotheses. The expectation was for three distinct factors to emerge: one describing declarative knowledge, one conceptual knowledge and one describing structural knowledge. The data actually supported two factors, with structural knowledge variables loading on one factor and declarative knowledge on the second factor. Concept knowledge loaded on both factors; this finding is discussed in the conclusions.

Introduction

A mental model is a knowledge structure, composed of concepts and the relations between them (Jonassen, Beissner & Yacci, 1993; Shavelson, 1974). Mental models are distinct from declarative and procedural knowledge (Jonassen & Tessmer, 1996). Anderson (1983) postulated the first stage of the development of expertise to be one of declarative encoding (distinct from procedural skills), where the subject is committing facts to memory. Declarative knowledge provides a base upon which subsequent learning can build relationships. The distinction between declarative knowledge and procedural skills is also made in R. Gagné's (1985) learning outcome taxonomy. Mental models, however, go beyond semantic relationships and skills acquisition. Mental models are knowledge structures that contain varied intellectual skills and knowledge. Jonnasen and Tessmer (1966) in addressing the distinctions, maintain:

> A mental model then contain three kinds of interconnected knowledge: knowing that (declarative), knowing how (skills), and knowing why (causal principles or functions). Conditional knowledge (knowing when) may also be a part of a mental model (p. 19).

(A glossary of terms associated with structural knowledge, construct validity and factor analysis is contained in Appendix A.)

There are at least three important implications of mental model research that should be considered by instructional designers. First, learners form mental models, whether the designer takes that fact into account or not, and inaccurate models can impede learning (Carroll & Thomas, 1982; Rouse & Morris, 1986; Norman, 1983). Second, troubleshooting can be facilitated through construction of mental models (Rouse & Morris, 1986; Gentner & Gentner, 1983). Similarly, structured knowledge has been found to be a powerful predictor of the ability to apply content knowledge (Gomez, Hadfield, & Housner, (1996). Finally, research on expert knowledge representation indicates that structural knowledge is essential to expert performance (Chi & Glaser, 1984; Larkin, McDermott, Simon, & Simon, 1980). Tardieu, Erlich, & Gyselinck, (1992), for example, found no difference between expert and novices at the prepositional level, but significant differences at the mental model level. Although the expert must have a sufficient knowledge base to draw upon, if it is not expertly structured, it is of little advantage.

Construct History

In recent years, several learning taxonomies have posited mental models as a learning outcome distinct from traditional outcomes of problem solving, concepts, and rules (Jonassen & Tessmer, 1996; Royer, Ciscero & Carlo, 1993). Before any construct becomes established as an independent learning outcome however, it should first be validated. As Rouse & Morris indicate, "...using a construct such as mental models results in requirements to define and illustrate the existence of such mechanisms (1986, p. 349)." While the mental models construct becomes increasingly important, it yet remains unvalidated. Kraiger & Wenzel (in press) have developed an elaborate

IR018459

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
M. Simonsen
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

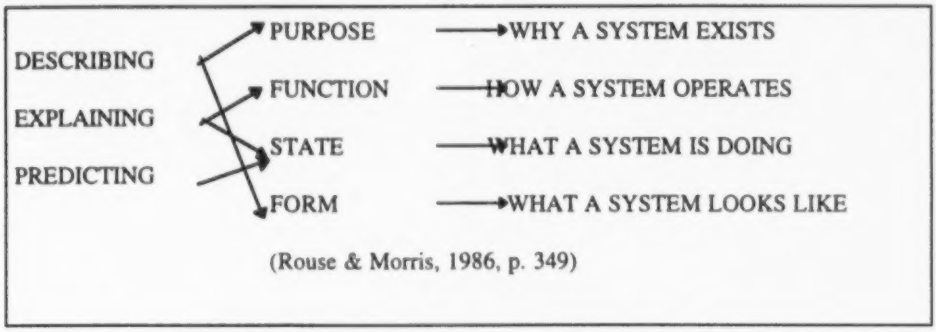
construct validity process for team mental models, but it is narrowly adapted to team performance, and cannot substitute for validation of the broader construct. Thus, the mental models construct remains unvalidated.

The evolution of the psychology of learning from a behaviorist to cognitive science has been in progress for at least several decades. The theoretical base in support of structural knowledge can be traced back to the early work of Bruner (Bruner, Goodnow, & Austin, 1956; Bruner, 1971); Ausubel (1960, 1968); Collins & Quillian (1969); Anderson (1974a; 1976; 1983; Anderson & Pirolli, 1984); and Rumelhart & Ortony (1977). Johnson-Laird reported that mental models as theoretical entities emerged as he wrestled with inference generation. He found the construct superior to other semantic representations for explaining meaning, comprehension and discourse (1983, p. 397). In general, these authors agree that memory structures are acquired by the learner. Mental models are a type of memory structure, distinct from skills, declarative or conceptual structures (Jonassen & Tessmer, in press).

Cognitive Importance

Building on the theory base that information in memory is encoded in and retrieved from a structure that preserves meaning, theorists posit mental models as powerful engines for higher order cognitive processes. The mental model enables the learner to solve problems, generate inferences, and make predictions about the system that is modeled (Johnson-Laird, 1981; Wilson & Rutherford, 1989). Figure 1 models some of those functions.

Figure 1. Purpose of Mental Models.



(Rouse & Morris, 1986, p. 349)

Rouse and Morris' work with troubleshooting and mental models would suggest another function of mental models learning: prediction what is wrong with a system (diagnosing).

Williams, Hollan, & Stevens (1983) offer a helpful operational definition of mental models. They see mental models as containing a powerful interpreter, capable of being used as an inference engine for conducting the functions depicted above, as well as delivering explanations and justifications, and serving as a mnemonic device for keeping the throttle within easy reach.

The Need for Construct Validity

In a recent special issue of *Educational Psychologist* on the Role of Knowledge in Learning and Instruction, guest editor Alexander (1996) considered the history of cognitive research. She lauded researchers who investigated prior knowledge, the organization and structure of knowledge and the differentiation of the novice learner and experts. However, she cites Pintrich (1994) in concluding that "Although the notion of knowledge was central in early cognitive research, the construct was not often systematically defined or clearly operationalized" (p. 91). Designers researched learning outcomes and utilized them in developing instruction, yet the outcomes may not have been sufficiently validated by construct validity research. In short, mental models constructs are not so much invalid as "unvalid". Their distinction from other outcomes has not been researched.

In a literature review, the authors found 202 hits on the ERIC database (1966-present) for "mental models," but the Boolean addition of "construct validity" to the term produced zero hits. The PsychLit database returned similar results with no construct validity articles. Perhaps more troubling was an ERIC search of "concept learning," a more recognized and traditional learning outcome. Of the 535 hits for that term, only one remained when "construct validity" was added. A search for "concept and construct validity" produced 105 hits, many of which

had to do with *self*-concept. The small number of construct validity studies dealt with only a few domains, notably Biology. In short, construct validity is conspicuously lacking in structured knowledge research.

Campbell & Fiske (1959) wrote the classic and enduring treatise on construct validity. It calls for convergent and discriminant validity of the construct across differing measures and distinctly different constructs in a multitrait-multimethod matrix (MTMM). Although most researchers today prefer factor analysis over the rigorous MTMM (Meier, 1994), the general principles of the matrix approach are timeless. In a review of construct validity three decades after its debut, Cronbach (1989) succinctly summarized its methodology:

> Instruments supposedly measuring the same variable should correlate...
> If constructs are distinct, their measures ought to rank persons differently, or in some other way to give distinctive reports (p. 153).

Messick (1980) describes construct validity as "a process of marshaling evidence" in support of the inferred meaning of the data set in the empirical relationships across measures. He uses the term trait validity to describe convergence of results on differing tests of a trait (construct) and nomological validity for the fit between theoretically divergent constructs and the data set (p. 1016). Cronbach and Meehl (1955) introduced the concept of a nomological (Greek, "legal") network. It would include "the theoretical framework" for the construct, an "empirical framework" for measuring it, and the relationships between the two frameworks (Trokim, 1996, p. 12).

The mental models validation task, then, is to measure learning outcomes with different instruments and demonstrate the reality of mental models by convergence and divergence of the measures of the constructs in accordance with a hypothesized theoretical framework. The reality of the mental models construct is difficult to validate, because theoretically a mental model may contain concepts or verbal information as its model components. Concepts may be a necessary component of a mental model; declarative knowledge may not. However, a series of studies using convergent and divergent validity will marshal the necessary evidence to establish the credence of the mental model construct as a distinct learning outcome. This study is a first step in that direction.

Purpose

This study investigated the convergent and discriminant validity of the mental models construct as indications of a distinct learning outcome. To accomplish this, the study was initiated to consider the relationship between mental models, declarative knowledge, concept learning, problem solving and troubleshooting performance. Although this was an exploratory, first look into the subject, hypotheses were formulated that the mental model "scores" will

1. factor load with concept learning scores (convergent validity with a metal models component);
2. not load with declarative knowledge measures (discriminant validity from a related outcome);
3. load with the dependent variable of problem solving performance (predictive validity of mental models acquisition); and that
4. the hypothesized structure of the exploratory factor analysis is for SPSS to identify three factors: one describing declarative knowledge, one conceptual knowledge and one describing structural knowledge.

Method

Subjects and Experts

The subjects were a class of junior-level Operations Management (MGT 396) students in a public university in the southeast. A prerequisite for the course was completion of an introductory course on spreadsheets. The class was made up of 25 students, 21 of whom were present on the day of testing: 9 females and 12 males. During MGT 396, students use spreadsheets as a tool in performing management tasks, and not as an end in themselves.

Three subject matter experts were employed to construct the testing instruments, the course instructor and two other college instructors. A Pathfinder network was drawn from each expert for comparison with the students' networks. The instructor was a full professor in the College of Business, and holds a Ph.D. in Engineering. The second expert taught business application software and introductory computer programming at the university, and

has an M.S. in Computer Science. A third expert taught business application software at a local junior college and has an MBA.

Design

The Delphi technique (Linstone & Turoff, 1975) was used to identify the concepts to be compared in a Pathfinder network and to generate a concept quiz. The technique was valuable because the experts teach very different student populations, and use spreadsheets for different instructional purposes (e.g., for production forecasting versus storage of financial data). Using the evolutionary Delphi process with a diverse base of expertise helped to identify concepts that epitomize spreadsheets. Sixteen core concepts were consequently selected by the experts.

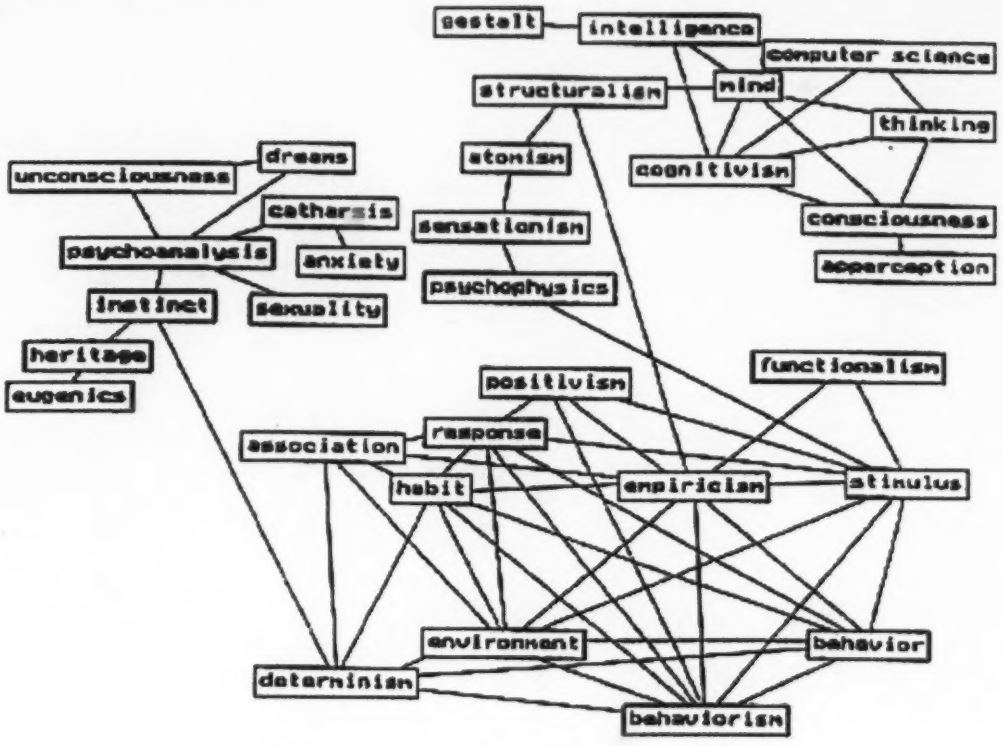
The final list of concepts was used as the terms for a Pathfinder network exercise using PCKNOT, an instrument used to assess structured learning attainment (Schvaneveldt, Durso, & Dearholt, 1989). Ten of the 16 concepts were also used in a concept learning quiz in which the students identified examples of the concepts from a list of 14 choices. A 20 question verbal information quiz was also developed from a spreadsheet text, and was modified by the experts. It consisted of multiple choice questions concerning spreadsheet terms, inputs and outputs.

Instruments

Pathfinder Associative Network. Pathfinder derives network structures from proximity data drawn from subject responses to pairwise comparisons of conceptual terms (Shvaneveldt, 1990). Pathfinder networks have been successfully used to capture changes in mental models as a function of learning (Gonzalvo, Cañas, Bajo, 1994; Cañas, Gonzalvo, & Bajo, 1992) and to capture novice-expert representational shifts (Kellogg & Breen, 1990; Goldsmith, Johnson, & Acton, 1991; Acton, Johnson, & Goldsmith, 1994; Schvaneveldt, Durso, Goldsmith, Bree, Cooke, & De Maio, 1985) through the use of similarity measures. In the present study, two instructors' networks served as the referent structures. The third network was rejected due to a lower internal coherence score.

The Figure 2 is an expert network constructed from concepts related to the history of psychology. It shows that the experts' network is highly organized, structured around several key concepts which form " node neighborhoods". (The length of the links connecting concepts is not part of the Pathfinder algorithm, and is not subject to interpretation.) The study that produced the network, like those above, used similarity as a measure of congruence between novices' and experts' structured knowledge.

Figure 2. Expert Pathfinder network.



Gonzalvo, Cañas, & Bajo, 1994, p. 611.

Coherence is another important statistic produced by Pathfinder networks and measures the consistency of the pairwise comparisons made by a subject. Coherence is a correlation between the proximity between a pair of items and the relatedness of those items with all others in the set-- the "interconnectedness of the knowledge representation" (Jonassen & Grabowsi, 1993). Higher correlations between the original proximity and the relatedness inferred from the indirect measures are indicative of expertise (Schvaneveldt, 1990). One might say that coherence is a plumb line held up against the subject's knowledge structure. Relatedness ratings for a given pair of concepts should "square" with the all of the other comparisons, individually and collectively.

Declarative knowledge quiz. A set of 22 concepts and a 15 multiple-choice questions were extracted from self-help questions in a spreadsheet textbook (Bidgoli, 1993). The textbook offered no reliability or validity coefficients for the questions, and none were compiled from the small data set at hand. Sample questions from the final 20-question quiz can be found in Appendix B.

Concept-matching quiz. The literature supports concept identification as a valid means of assessing concept knowledge (Gagné, Briggs, & Wager, 1974). One method of doing so, is in a matching format. The subjects were asked to match 10 spreadsheet concepts to a list of 14 choices. An evaluation of this assessment will be found under the Discussion section. Again, this is an unvalididated instrument. The matching quiz can be found in Appendix C.

Variables

The variables included in a factor analysis must be representative of attributes of the construct(s) under consideration. It is tempting to include a large number of variables and observe what emerges from the analysis. There are at least three reasons for not doing so. First, as discussed later, supportable factor identification hinges on a reasoned theoretical basis for the construct. Second, a parsimonious set of variables will decrease likelihood of chance significance being observed. For example, a study with 30 variables will produce 435 correlations. At a significance level of .05, there could be 20 relationships deemed significant just by chance (Hair, Anderson, Tatham,

BEST COPY AVAILABLE

& Black, 1995). Finally, since statisticians recommend a subject-to-variable ratio between five- and ten-to-one (Hair et al., Crocker, 1986), the number of subjects needed for a study can be kept manageable, usually a consideration in social science contexts.

Since the object of factor analysis is to reduce some set of variables to a smaller number of factors which adequately represent the original set (see for example, Kim & Mueller, 1978), the variables must have a substantial degree of covariation, known as communality. The variables in this study exhibit a moderately high degree of covariance. The variables invested were:

1) the results of the verbal information quiz;
2) results of the concept matching quiz;
3) internal coherence of the students' Pathfinder networks;
4) similarity between students' pathfinder networks and their instructor's network; and
5) similarity with a second instructor's network.

Exploratory Factor Analysis

Factor analysis is a multivariate statistical method used to uncover the underlying structure in a data set. This is accomplished by identifying a small number of factors (latent variables) that are representative of the relationships in a larger set of constituent variables. The question is not one of relationships between dependent and independent variables, but of identifying factors that are not directly observable from a set of variables. The distinguishing feature of factor analysis is the assumption that the covariation amongst the variables is due to an underlying structure of common factors (Kim & Mueller, 1978). The study sought to determine whether the knowledge variables will cluster around factors that suggest the existence and measurability of mental models.

In confirmatory factor analysis, we enter the analysis with an *a priori* structure for the data. CFA seeks to answer the question, "Do the data fit the hypothesized model?" and acts as a deductive reasoning tool for theory testing. Although the authors had formulated hypotheses about emergent factors and relationships among the variables, this was very much an exploratory factor analysis. In EFA, the researcher allows the latent structure to emerge from the data set a *posteriori*. EFA serves as an inductive reasoning tool for theory building (Bryant & Yarnold, 1995, p. 109). The number of expected factors was not imposed upon the factor analysis. Rather, Eigenvalues of 1.0 or greater (latent root criterion), and a Scree plot were used. An Eigenvalue (the column sum of squared loadings for a factor) represents the amount of variance of the original set of variables accounted for by a given factor. The Scree test criterion seeks to identify the point at which unique variance begins to exceed the common variance in the structure. Latent roots are plotted against the number of factors. The point at which the slope begins to straighten out represents to maximum number of factors to extract (Hair et al, 1995). Four steps were used in the factor analysis:

1. compute the correlation matrix for the variables;
2. extract the factors from the correlation matrix;
3. rotate the factors to make them more interpretable; and
4. analyze and name the factors.

In factor rotation, the reference axes of the factors are rotated, seeking a theoretically more meaningful representation of the original variables. Varimax, an orthogonal rotation procedure, was used in the study. Orthogonal rotation maintains the factors at a 90 degree orientation; that is, the factors are uncorrelated with each other. That is reasonable, in that the study seeks to identify unique factors that represent distinct constructs. Figures 3 and 4 are an oversimplification, but graphically represent the theory and process of rotation.

Figure 3. Unrotated variables and factors.

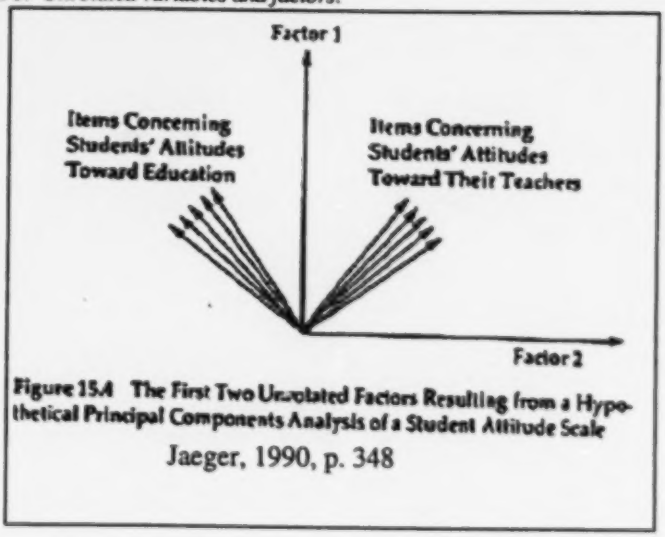

Factor 1
Items Concerning Students' Attitudes Toward Education
Items Concerning Students' Attitudes Toward Their Teachers
Factor 2

Figure 15.4 The First Two Unrotated Factors Resulting from a Hypothetical Principal Components Analysis of a Student Attitude Scale

Jaeger, 1990, p. 348

Figure 4. Rotated variables and factors.

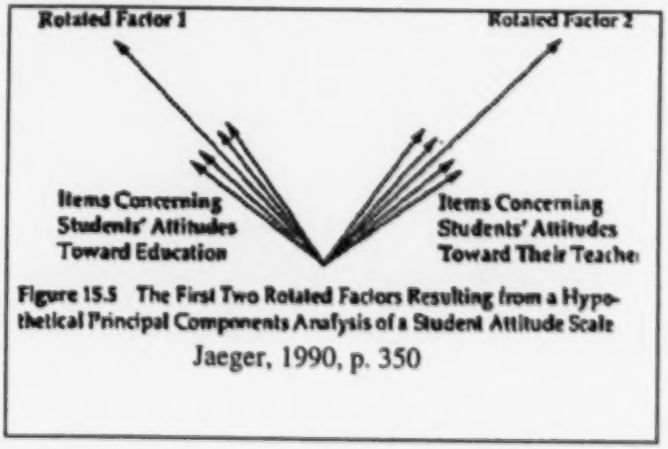

Rotated Factor 1
Rotated Factor 2
Items Concerning Students' Attitudes Toward Education
Items Concerning Students' Attitudes Toward Their Teache

Figure 15.5 The First Two Rotated Factors Resulting from a Hypothetical Principal Components Analysis of a Student Attitude Scale

Jaeger, 1990, p. 350

"Simple structure" would describe a factor matrix loading pattern where each variable loaded very heavily on one factor in the matrix, and very little or not at all on the others. An example might look like the factor matrix in Table 1.

Table 1. Example of simple structure.

	Factor 1	Factor 2	Factor 3
Course GPA	1.0	0	0
Problem Solving Lab	0	0	1.0
Pathfinder Coherence	0	0	1.0
Pathfinder Similarity	0	0	1.0
Declarative Knowledge Test	1.0	0	0
Concept Test	0	1.0	0

In an imperfect analogy, the factor loadings can be thought of as correlation coefficients for the respective variables (Jaeger, 1990). That is, the results of the Declarative Knowledge Test would have perfect correlation with Factor 1, and no relationship with the other two factors. In this study, we expect to find variable loadings that would indicate the reality of mental models. If, for example, Verbal Information in our study were to load heavily on the same factor as Pathfinder Internal Coherence, the existence of mental models would not be supported.

It is the factor matrix produced in step three that must be analyzed in conjunction with the theory base for the constructs. The first three steps were simply a procedural function of the software program. The last step incorporates both art and science: a theory-based interpretation of the statistical results.

Results

Step 1: The Correlation Matrix. Table 2 present the correlation coefficients. An explanation of the variables may be found in the factor matrix under Measure.

Table 2. Pearson correlations among the variables.

	COHERENCE	CONCEPTS	VERBINFO	ZSMLRTY	PSMLRTY
COHERENCE	---				
CONCEPTS	.222	---			
VERBINFO	-.007	. 511**	---		
ZSMLRTY	.619***	.391*	.051	---	
PSMLRTY	.518**	.164	.405		.617***

—

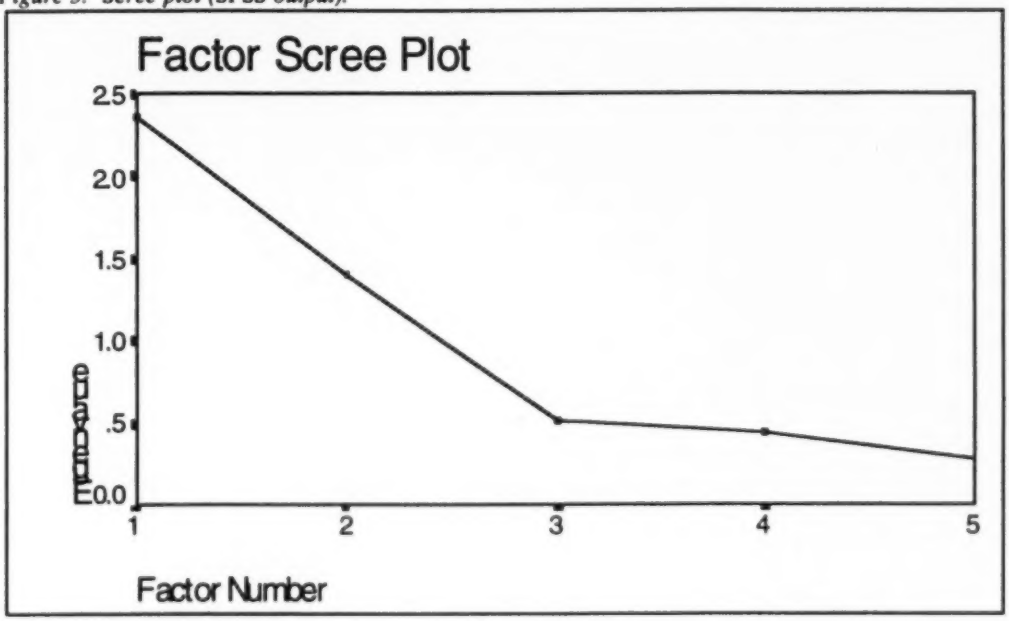
n=14; *Correlation (1-tailed) is significant at the 0.10 level; ** the 0.05 level ; *** the 0.01 level .

Verbal information (declarative knowledge) did not correlate with any of the mental models variables, but moderately correlated with concept identification (r=.51; p=.03). Concept identification was moderately correlated with student-course instructor similarity (r=.39), and was statistically significant at the .10 level. This is an acceptable alpha level for exploratory studies (Borg, & Gall, 1989). Coherence was moderately and significantly

correlated with similarity (r=.62; p=.01). The correlation between coherence and concepts was .222; with verbal information, -.01. Neither relationship was statistically significant.

Step 2: Factor Extraction. Two methods were used to determine the number of factors to extract: Eigenvalues of 1.0 or greater (Appendix D), and a Scree plot. The knee of the Scree plot (Figure 5) falls at the three factor point, but the third factor accounts for a relatively small amount of the total variance, making it of little use as factor in the study.

Figure 5. Scree plot (SPSS output).



Step 3: Factor Rotation. Varimax rotation with Kaiser normalization was used, and converged in three iterations. Two factors were extracted. The factor loading matrix is presented in Table 3.

Table 3. Factor loading matrix.

	Factor 1	Factor 2	% Var	Measures
1 COHERENCE	.833	.034	47.2	1 Internal consistency of Pathfinder models.
2 CONCEPTS	.270	.839	27.9	2 Score on a concept quiz.
3 PSMLRTY	.829	.000	10.3	3 Similarity with a second instructor's Pathfinder network.
4 VERBINFO	-096	.888	09.0	4 Score on verbal info quiz
5 ZSMLRT	.871	.191	05.7	5 Similarity with course instructor's Pathfinder network.

Step 4: Analysis and Identification. The factor matrix above approaches simple structure, with each variable loading high on one factor and low on the other, with the exception of the scores on the concept knowledge quiz. That variable, though loading highly with verbal information on Factor 2, also loads moderately with the more structural variables on Factor 1 (Coherence, Zsimilarity, and Psimilarity). Coherence accounted for 47.2 percent of the total variance amongst the variables; scores on the concept quiz accounted for almost 28 percent. Communality (the amount of variance of a given variable shares with the other variables) was moderate to high, ranging from .69 to .80, with a mean of .75 (Appendix D). That is, about 75 percent of the variance in the five variables selected is held in common; this is a fertile environment for factor analysis. Using Varimax rotation, SPSS extracted two factors. Based on the mental models literature, emergent Factor 1 is tentatively labeled Mental Models, and is composed of network coherence and similarity measures. Factor 2 is tentatively labeled Semantic Knowledge, and is comprised of verbal information and concept measures.

Discussion

The extracted factors accounted for about 75 percent of the variance within each of the variables, indicating that the two-factor model was a good fit for the five variables used. Some of the results supported the research and hypotheses, some did not. The most important finding was that all three purely structural variables (coherence and both similarity variables) loaded almost exclusively on one factor and verbal (declarative) knowledge loaded exclusively on the other. These results support the discriminant validity of the mental models construct. Moreover, network coherence (the structural "plumb line") accounted for almost half of the total variance amongst the variables.

As expected, the students' mental model internal coherence correlated much more positively with concepts than with verbal information, although the correlation was not large. This supports the proposition that mental models are comprised of conceptual relationships more than declarative knowledge (Jonnasen & Tessmer, 1996). However, verbal information correlated positively with concept identification, and the latter loaded more heavily on the "semantic knowledge" factor, indicating that these outcomes are not as discrete as instructional design literature hold them to be. Part of the overlap can be explained by prerequisite knowledge: verbal information of concept attributes can facilitate concept learning. It is interesting that in the unrotated factor matrix (Appendix D), Concepts loaded .57 on Factor 1 and .68 on Factor 2. This is obviously a much more even spread of its variance than the .27 to .84 ratio in the rotated matrix. The loading of the concept learning variable when a rigorous test of concept understanding is used will be instructive.

It was noted above that the coherence of the students' Pathfinder networks accounted for nearly half of the total variance amongst the knowledge variables. Some discussion of the distribution of the remaining variance is in order. The subjects were attending their third class meeting of Operational Management. Their instructor had a rich, well developed knowledge base in the use of spreadsheets in novel and innovative contexts. Each student had taken a prerequisite course, where spreadsheets were presented as the course content in a simple, mechanistic manner. The second instructor taught such an introductory course in spreadsheets. The second instructor's approach to spreadsheets was more "in line" with the students' experience with spreadsheets. Had the study continued, the course instructor's Pathfinder network would be hypothesized to correlate increasingly more with the students'.

The concept learning outcome unexpectedly loaded with the "semantic" factor more heavily than the 'structured" factor. Using a matching exercise for the identification of concepts was apparently not a rigorous enough test. Memorized information would have been sufficient to correctly identify the concepts. The quiz did not require a deep understanding of the concepts on the part of the subjects. Tessmer, Wilson & Driscoll (1990) offer alternative recommendations for the measurement of concept learning that go beyond simple example identification:

1) using the concept in conversation, writing and argumentation (Brown et al., 1989);
2) simulating or role-playing the concept (Tessmer, Jonassen & Caverly, 1989); and
3) making judgments or criticisms on the basis of the concept (p. 48).

It is anticipated that if a more demanding measurement instrument was used to assess concept knowledge, the results would load on the "structural knowledge" factor, and not correlate as strongly with verbal (declarative) knowledge.

This initial investigation has several limitations. The study was terminated after the data set considered in this paper was collected, which caused some of the limitations. Others have to do with the instruments and methodology used. The limitations are :

1. Subject mortality. Five students chose the same level of comparison for each pair of concepts during the Pathfinder exercise. There was no payoff for them to take the time to make the large number of comparisons. Pairwise comparison measures can be tedious tasks that tempt the subject to hasten their comparison ratings. To the

degree that there was a correlation between the population that failed to perform the comparisons and any of the variables, bias was introduced.

2. Generalizability. Because neither random selection or assignment was used, the external validity is in question.

3. Missing variables. Because the study was terminated early, potentially important variables were missing from the matrix. When factored in, they may have had significant impact on the resultant factor matrix.

4. Insufficient n. The number of subjects (14) was insufficient for a study of this nature.

5. Unvalidated instruments. The verbal information and concept tests were unvalidated. Without a valid instrument, it is difficult to maintain that a hypothesized construct has actually been measured.

6. Convergent validity. Only Pathfinder networks were used to assess structured knowledge.

Recommendations

Some final recommendations should be considered for future research into construct validity for mental models. First, provide an incentive for subject participation in the Pathfinder data collection. Next, use an *n* equal to five-to-ten times the number of variables being factored (Crocker, 1986), ideally 50 in this study. Third, test concept knowledge by using rigorous measurement of inferential abilities of the students. Fourth, use only valid instruments to measure the variables. Finally, use multiple methods and measures. Perhaps card sorts and interviews with students in addition to Pathfinder networks to assess mental model attainment and similarity with experts. The alternative methods for concept knowledge assessment mentioned earlier could be used. Also, multiple regression could be used to measure the predictive power of mental models upon problem solving.

Conclusions

Instructional designers have long been involved in the research and use of learning outcomes, but have not always been as quick to validate those outcomes before incorporating them into taxonomies. This study was a first step towards construct validation of the mental models learning outcome. A great deal is left to be done.

Cronbach's (1984) view is that "Construct validation is a fluid, creative process...to develop an interpretation, persuade others of its soundness, and revise it as inadequacies are recognized (p. 149)." We are still in that creative process, but we have a tentative interpretation of the mental models construct. Now we need to move toward "soundness" through instruments and variables that have persuasive power. We close with Messick's (1980) summary:

> Thus, the paradox that measures are needed to define constructs and constructs are needed to build measures is resolved, like all existential dilemmas in science, by a process of successive approximation (p. 1016).

Future studies will continue the process, and move successively closer to a valid mental models construct.

Appendix A

Definitions

Structured Knowledge Terms

Structural knowledge describes an individual's organization of ideas (knowledge structure) about different content domains. This knowledge is essential to understanding the content and the ability to apply it (Jonassen & Grabowski, 1993, p. 434).

Propositions represent the atomic units of meaning and can be used to represent the meaning of sentences and pictures. The interconnections among propositions define a *network* (Anderson, 1990, p. 143).

Certain groups of propositions group together in larger-order units called *schemas.* Here, the abstraction is from specific instances to generalizations about the category from which these instances come (Anderson, 1990, pp. 144, 133).

Mental models are the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states (Rouse & Morris, 1986, p. 351).

Pathfinder networks represent concepts as nodes and distances between concepts as links in a network. The Pathfinder algorithm takes proximity values as input and produces as output a network with shortest-path link distances (Acton, Johnson, & Goldsmith, 1994, p. 305).

Construct Validity Terms

The term *construct* is used...to refer to something that is not observable, but is literally *constructed* by the investigator to summarize or account for the regularities or relationships that he observes in behavior (Thorndike & Hagan, 1986, p. 70).

Construct validity refers to the nature of the psychological construct or characteristic being measured by the instrument. How well does this construct explain differences in the behavior of individuals or their performance on certain tasks (Fraenkel & Wallen, 1996, p. 154).

Factor Analysis Terms

Communality is the amount of variance an original variable shares with all other variables included in the analysis.

An *Eigenvalue* (or latent root) represents the amount of variance accounted for by a factor.

A *factor* is a linear combination (variate) of the original variables. Factors also represent the underlying dimensions (constructs) that summarize or account for the original set of observed variables.

Factor loadings are correlations between the original variables and the factors, and the key to understanding the nature of a particular factor.

Factor rotation is the process of manipulating or adjusting the factor axes to achieve a simpler and pragmatically more meaningful factor solution (Hair, Anderson, Tatham, & Black, 1995, pp. 365-366)

Appendix B

Sample Questions -- Declarative Knowledge Quiz

Select the most correct answer for each question by writing the corresponding letter.

_____1. The intersection of a row and column is called a
a. cell
b. pointer
c. function
d. formula
e. none of the above

_____2. Using a spreadsheet for what-if analysis, you can
a. change one value and see the result on the rest of the worksheet
b. change one value and create a new graph
c. do both a and b
d. change one graph and create a new value
e. do none of the above

_____12. Compared to /Range Format, /Worksheet Global Format has
a. lower priority
b. higher priority
c. equal priority
d. priority determined by the type of calculation
e. none of the above

_____15. The most suitable graph for time series analysis is a (an)
a. pie chart
b. exploded pie chart
c. any graph that uses symbols
d. line graph
e. none of the above

_____16. The X range is used in
a. pie charts
b. exploded pie charts
c. graphs that use symbols
d. line graphs
e. all graphs

Apppendix C

Concept Matching Quiz

Match the concept in Column A with the corresponding example in Column B. Examples may be used once, more than once or not at all.

COLUMN A	COLUMN B
_____21. Argument	A. AB2-CD4
_____22. Cell Address	B. ACME.WK1
_____23. File	C. A5
_____24. Formula	D. {CALC}
_____25. Graphics	E. B1..C4
_____26. Label	F. Champions.DOC
_____27. Macro	G. FEMALE
_____28. Numeric Constant	H. +B1*H1/2
_____29. Range	I. WYSIWYG
_____30. What-if	J. 3.1416
	K. @MaxA1..B3
	L. @Length("Title"&"Sub-Title")
	M. .PIC
	N. 9.00

Thanks for your help with this project.

Appendix D

SPSS Factor Analysis Outputs

Correlation Matrix:

	COHERE	CONCEPTS	PSMLRTY	VERBINFO	ZSMLRTY
COHERE	1.00000				
CONCEPTS	.22243	1.00000			
PSMLRTY	.51848	.16370	1.00000		
VERBINFO	-.00733	.51050	.00653	1.00000	
ZSMLRTY	.61867	.39096	.61673	.05086	1.00000

Determinant of Correlation Matrix = .2165431

Principal Components Analysis (PC) **Initial Statistics:**

Variable	Communality	Factor	Eigenvalue	Pct of Var	Cum Pct
COHERE	1.00000	1	2.35804	47.2	47.2
CONCEPTS	1.00000	2	1.39304	27.9	75.0
PSMLRTY	1.00000	3	.51319	10.3	85.3
VERBINFO	1.00000	4	.45074	9.0	94.3
ZSMLRTY	1.00000	5	.28499	5.7	100.0

Unrotated Factor Matrix

	Factor 1	Factor 2
COHERE	.78480	-.28197
CONCEPTS	.56617	.67579
PSMLRTY	.76798	-.31214
VERBINFO	.24518	.85838
ZSMLRTY	.87845	-.15034

~~VARIMAX rotation, Kaiser Normalization.~~

Rotated Factor Matrix:

	Factor 1	Factor 2
COHERE	.83323	.03386
CONCEPTS	.27047	.83910
PSMLRTY	.82899	-.00042
VERBINFO	-.09561	.88757
ZSMLRTY	.87051	.19105

VARIMAX converged in 3 iterations.

Final Statistics:

Variable	Communality	Factor	Eigenvalue	Pct of Var	Cum Pct
COHERE	69541	1	2.35804	47.2	47.2
CONCEPTS	.77724	2	1.39304	27.9	75.0
PSMLRTY	.68722				
VERBINFO	.79692				
ZSMLRTY	.79428				

References

Acton, W. H., Johnson, J. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. Journal of Educational Psychology, 86(2), 303-311.

Alexander, P. A. (1996). The past, present, and future of knowledge research: A reexamination of the role of knowledge in learning and instruction. Educational Psychologist, 31(2), 89-92.

Anderson, J. R. (1974a). retrieval of prepositional information from long-term memory. Cognitive Psychology, 6, 451-474.

Anderson, J. R. (1976). Language, memory, and thought. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. R., & Pirolli, P. L. (1984). Spread of activation. Journal of Experimental Psychology, 10(4), 791-799.

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology, 51(5), 267-272.

Ausubel, D. P. (1968). Educational psychology: A cognitive view. New York: Holt, Rinehart, and Winston.

Bidgoli, H. (1993). Information systems literacy: Lotus 1-2-3 (Release 2.3). New York: Macmillan.

Borg, W. R. & Gall, M. D. (1989). Exploring relationships between variables: The causal-comparative method. In Educational research: an introduction (5th ed.). New York: Longman.

Bruner, J. S. (1971). The relevance of education. New York: Norton.

Bruner, J. S., Goodnow, J., & Austin, G. A. (1956). A study of thinking. New York: Wiley.

Bryant, F. B. & Yarnold, P. R. (1995). Principle components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics. Washington, DC: American Psychological Association.

Campbell, D. T., & Fiske, D. W. (1959, March). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Cañas, J. J., Gonzalvo, P., & Bajo, M.-T. (1992, September). Mental models and computer programming. Paper presented at the fifth conference of the European Society for Cognitive Psychology, Paris, France.

Carroll, J. M. & Thomas, J. C. (1982). Metaphor and the cognitive representation of computing systems. IEEE Transactions on Systems, Man, and Cybernetics, SMC12, 107-116.

Chi, M. T., & Glaser, R. (1984). Problem solving abilities. In W. G. Chase (Ed.), Human abilities: An information processing approach (pp. 227-248). San Francisco: Freeman.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Crocker, L. M. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston.

Cronbach, L. J. ((1984). Essentials of psychological testing (4th ed.). New York: Harper & Row.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

Gagné, R. M. (1985). The conditions of learning (4th ed.) New York: Holt, Rinehart, and Winston.

Gagné, R. M., Briggs, L. J., & Wager, W. W. (1974). Principles of instructional design. New York: Holt, Rinehart, and Winston.

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), Mental models (pp. 99-129). Hillsdale, NJ: Erlbaum.

Gomez, R. L., Hadfield, O. D., & Housner, L. D. (1996). Conceptual maps and simulated teaching episodes as indicators of competence in teaching elementary mathematics. Journal of Educational Psychology, 88(3), 572-585.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. Journal of Educational Psychology, 83, 88-96.

Gonzalvo,P., Cañas, J. J., & Bajo, M.-T. (1994). Structural representations in knowledge acquisition. Journal of Educational Psychology, 86(4), 601-616.

BEST COPY AVAILABLE

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). Multivariate data analysis with readings (4th ed.). New York: Macmillan.

Jaeger, R. (1990). Statisics as a spectator sport. Newbury Park, CA: Sage Publications.

Johnson-Laird, P. (1983). Mental models. Cambridge, MA: Harvard University Press.

Jonassen, D. H., Beissner, K., Y Yacci, M. A. (1993). Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jonassen, D., & Tessmer, M. (1996). An outcomes-based taxonomy for the design, development, and research of instructional systems. Training Research Journal, 2, 1-41.

Kellogg, W. A., 7 Breen, T. J. (1990). Using Pathfinder to evaluate user and system models. In R. W. Schvaneveldt (Ed.), Pathfinder associative networks: Studies in knowledge organization (pp. 179-195). Norwood, NJ: Ablex.

Kim, J. & Mueller, C. W. (1978). Introduction to factor analysis: What it is and how to do it. (Vol. Ed.). In J. L. Sullivan & R. G. Niemei (Series Eds.). Quantitative applications in the social sciences. Beverly Hills, CA: Sage University Press.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation.

Kraiger, K., & Wenzel, L. H. (in press). A framework for understanding and measuring shared mental models of team performance and team effectiveness. In E. Salas, & C. Prince (Eds.), Team performance assessment and measurement: Theory, methods and applications.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

Linstone, H. A., & Turoff, M. (Eds.). (1975). The Delphi method: Techniques and applications. Reading, MA: Addison-Wesley.

Meier, S. T. (1994). The chronic crisis in psychological measurement and assessment. San Diego: Academic Press.

Messick, S. (1980, November). Test validity and the ethics of assessment. American Psychologist, 35(11), 1012-1027.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), Mental models (pp. 7-14). Hillsdale, NJ: Erlbaum.

Qualls, A. L., & Moss, A. D. (1996, April). The degree of congruence between test standards and test documentation within journal publications. Educational and Psychology Measurement, 56(2), 209-214.

Rouse, W. B. & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100(3), 349-363.

Royer, J. M., Ciscero, C. A., & Carlo, M. S. (1993). Techniques and procedures for assessing cognitive skills. Review of Educational Research, 63(2), 201-243.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory: In R. C. Anderson, J. R. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Erlbaum.

Schvaneveldt, R. W. (Ed.). (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 24, pp. 249-284). San Diego: Academic Press.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Bree, T. B., Cooke, N. M., & De Maio, J. C. (1985). Measuring the structure of expertise. International Journal of Man-Machine Studies, 23, 699-728.

Shavelson, R. J. (1972). Some aspects of correspondence between content structure and cognitive structure in physics instruction. Journal of Educational Psychology, 63, 225-234.

Shavelson, R. J. (1974). Methods for examining representation of a subject-matter structure in a student's memory. Journal of Research in Science Teaching, 11, 231-249.

Tardieu, H., Ehrlich, M.-F., & Gyselinck, V. (1992). Levels of representation and domain-specific knowledge in comprehension of scientific texts. Language and Cognitive Processes, 7(3/4), 335-351.

Tessmer, M., Wilson, B, & Driscoll, M. (1990). A new model of concept teaching and learning. Educational Technology Research & Development, 38(1), 45-53.

Trokim, W. M. K. (1996). Construct validity. [On-line.] Available URL: http://trochim.human.cornell.edu/kb/constval.htm.

Williams, M. D., Hollan, J. D., & Stevens, A. L. (1983). Human reasoning about a simple physical system. In D. Gentner & A. L. Stevens (Eds.), Mental models (pp. 131-153). Hillsdale, NJ: Erlbaum.

Wilson, J. R. & Rutherford, A. (1989). Mental models: Theory and application in human factors. Human Factors, 31,(6), 617-634.